Itemized Response

Reg. To: US Patent Application # 10/662,552

Applicant: Zhang, Xiaohui, 4400 E Broadway blv, Tucson, AZ 85711 Fax: 520-202-3340 ; Phone: 520-202-3333
Title: Methods And System For Bio-Intelligence From Over-The-Counter Pharmaceutical Sales
Mail Date: March 16, 2007
Examiner: Mr. Neal Sereboff, Art Unit 3626

Note of Response:

Reference U is *Goldenberg et al, Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales,* Proc Natl Acad Sci U S A. 2002 April 16;

Reference V is Armstrong et al, ***Updated Guidelines for Evaluating Public Health Surveillance Systems,*** *CDC* July 27, 2001 / 50(RR13);1-35

Reference [3] is the attached reference by X. Zhang et all, ***A Biointelligence System for Identifying Potential Disease Outbreaks,*** IEEE Engineering in Medicine and Biology (EMB), Vol 23, Number 1, 2004.

Response

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
1	Claim 1 -15 are pending	Response within 3 months
2	Advise to have a patent attorney	Applicant will have a patent attorney soon.
3	Claim 6 and Claim 9 in improper form because of multiple dependent claim should be in the alternative	Corrected Claim 6 and Claim 9 to the alternative form, following the examples in Section 7.45 of MPEP 608.01(n)
4	Appropriate correction is required for verb	Corrected verb 'are' to 'is'
5	Claim 11 partially repeated the language of Claim 10, Appropriate correction is required	Removed the repeated sentence.

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
6	Amend the claim, since Claim 15 dependent upon Claim 1, not Claim 12	Removed the sentence that is dependent on Claim 12
7	Copy of '35 U.S.C. 101	Noted
8	'non-statutory subject matter'	The claims 1-15 represent a new useful system, its application demonstrated that it is statutory subject matter according to 35 U.S.C. 101. For elaboration please review the response letter, the attached reference [3] and responses below.
9	(1) Question of 'the results be reproducible' (2) Question on 'Subjective component'	(1) The system and the method in claim 1- 15 have been implemented in a computer system, and said system has been helping public health workers in identifying disease outbreaks. The results are completely reproducible for the given data and conditions. Please see the attached reference publication by X. Zhang et al: A Biointelligence System for Identifying Potential Disease Outbreaks, in IEEE Engineering in Medicine and Biology (EMB), Vol 23, Number 1, 2004. (2) There is no 'subjective component' in this system; it intends to overcome the limits from 'subjective components'. The results are completely based on an analysis of the data and are not dependent on subjective feelings of the analyst. The rules are derived directly from the investigated data itself not from "experience", or other subjective sources. The system automatically takes the historical data to derive the categorized public health status, reference values, decide the state transitions, and make decisions, in a specified place.
10	Implementation of the rule system	(1) The system has been implemented in a computer system,

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		including the rule systems. Several of the included figures are taken from sample implementations. Please see the attached reference publication by X. Zhang et al: A Biointelligence System for Identifying Potential Disease Outbreaks, in IEEE Engineering in Medicine and Biology (EMB), Vol 23, Number 1, 2004. (2) In the patent application, the rule systems are demonstrated in detail (please see paragraph 86 to paragraph 93), step by step, the rule system decides the state transition and inputs/outputs with data referenced in fig. (5) to (8). In fact, the same examples were provided to two different software developers, and they both implemented their application systems independently, one system was published by X. Zhang et al: A Biointelligence System for Identifying Potential Disease Outbreaks, in IEEE Engineering in Medicine and Biology (EMB), Vol 23, Number 1, 2004. The other system is currently in
11	Copy of first paragraph of 35 U.S.C. 112 "The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention."	Noted

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
12	Question (1) how the invention overcomes existing problems (Background of Invention)	(1) Existing system consider a categorized public health status in a community (such as flu situation) as one of two status: 'normal' or 'outbreak'. Traditionally, analyses are conducted in such binary mode. This is a shortcoming that results from the use of traditional statistical approaches. The usual objective is to try to detect an 'outbreak' (one out of the two status) early. However, the system presented in claim 1- 15 we model for the first time a dynamic change of a categorized public health status in a community by a set of state variables (seven state variables) and the transition of state variables. The seven state variables defined by the applicant are the minimum set to characterize a complete process of a community's public health status. This approach is inherently different from statistical approaches as it integrates dynamic model, rule systems and information technique.
	(2) lack of method of practicing and working example	(2) The presented system has been described in sufficient detail as a blue print for software engineers to implement as evident in the two independent and successful implementations, which both include the outlined rule systems. Please see the attached reference publication by X. Zhang et al: A Biointelligence System for Identifying Potential Disease Outbreaks, in IEEE Engineering in Medicine and Biology (EMB), Vol 23, Number 1, 2004.

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
13	'claim 1 – 15 are rejected, based on a disclosure is not enabling' 'structural implementation'	The system is implemented and already deployed; the implemented system has been running since 2003 and has successfully aided a Public Health department in the cognition of several otherwise unknown disease outbreak events. Please see the attached reference publication by X. Zhang et al: A Biointelligence System for Identifying Potential Disease Outbreaks, in IEEE Engineering in Medicine and Biology (EMB), Vol 23, Number 1, 2004. Examples of implementation are provided in the attached publication by X. Zhang et al: A Biointelligence System for Identifying Potential Disease Outbreaks, in IEEE Engineering in Medicine and Biology (EMB), Vol 23, Number 1, 2004.
14	Claim 1-15 are rejected, as failing to comply with enablement requirement. working example or instruction	Working example is provided in the attached publication by X. Zhang et al: A Biointelligence System for Identifying Potential Disease Outbreaks, in IEEE Engineering in Medicine and Biology (EMB), Vol 23, Number 1, 2004.
15	quotation of the second paragraph of 35 U.S.C. 112: "The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention."	Noted Please see response to Action 16.

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
16	Need to define the invention required by 15	Two of obvious inventions are identified as examples: It is the first time (an invention), that a dynamic change of a categorized public health status in a community is modeled by a set of state variables (seven variables) and the transition of state variable in a dynamic model. The seven state variables defined by the applicant are the minimum set to characterize a complete process of a community's categorized public health status. It is the first time, which in dynamic modeling in state space approach, the state transition matrix is systematically defined by a combination of numerical functions and rule base and knowledge management.
17	Language and format The claim must be in one sentence form only	 Modified format.
18	Use of 'the' and 'a' in Claim 1	Corrected 'the rule system' to 'a rule system' in Claim 1
19	Use of 'the' and 'a' in Claim 1	Corrected
20	Use of 'the apparatus'	Corrected in Claim 1
21	Use of 'the' baseline	Corrected to 'a'
22	Claim 3 -5 failing to point out and distinctly claim the invention Claim 3 -5 dependent upon Claim 2	 Modified Claim 3 – 5.
23	Claim 10 failing to point out and distinctly claim the invention	Mended the sentence, as

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
	Question: what is being mapped	The structural components are mapped incorporating their confidence levels.
24	quotation of the paragraph of 35 U.S.C. 102: "A person shall be entitled to a patent unless – (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of the application for patent in the United States."	Noted.
25	Claim 1 – 5, 7, 10-15 are rejected, because of Goldenberg et al "Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales" , reference U	Please see the response to Action 26.
26	As per claim 1, The first paragraph: 'Goldenberg teaches the system for detecting an unusual public health status and for modeling the change of categorized public health status from OTC pharmaceutical data'	In reference U, Goldenberg, as all other public health workers, does not define 'public health status' as a set of state variables, does not model 'the change of categorized public health status', there is no word of 'public health status' appeared in Reference U. Goldenberg tries to use OTC data to detect an outbreak only (distinguish an 'outbreak' from 'normal'), he doe not mention how to model the detection in the space of state variables as described in Claim 1. Goldenberg's approach is the traditional statistical approach with a signal detection for Normal or Attack (outbreak); while the system presented in claim 1- 15 describes a complete process for a categorized public health status in a community.

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		Conclusion: Claim 1 is different from the one in the reference.
	The second paragraph: A measurement scheme defined by a set of variables and calculation of categorized daily OTC sale data in a specified geographical scale (page 2, Tracking Grocery Data section, paragraph 1 where geographical scale is noted by address)	Quotation of Tracking Grocery Data section, paragraph 1 (of reference U): "**Tracking Grocery Data** Grocery and OTC medication sales have three main advantages for the detection of an outbreak: First, these datasets are typically very large and rich, including information on each purchased item and in many cases include customer information(e.g., address). They are also available on a more frequent scale, such as daily and even hourly basis, and do not include delays in reporting as compared with medical and public health sources which are typically collected weekly or even less frequently, and might contain delays. Second, the outbreak footprint would probably exist in these data earlier than in medical or public health data, because of self treatment that people usually pursue before seeking medical assistance. Third, although grocery and OTC sales do not measure illness directly, we might infer specific symptoms experienced by purchasers at a relatively early stage of the onset of the disease." Analysis: In this paragraph, Goldenberg summarized the advantages of using OTC data. We agree with his analysis of the benefits, which is why we are working with the same data source. No specific method is presented. Goldenberg mentioned the use of customer information (e.g. address). Immediately in the following paragraph, Goldenberg pointed out 'the main problem using grocery and OTC medication sales is their noise nature'. In Claim 1 – 15, no a customer address is required or used. In section of Summary of Invention, a study area or a geographical level is defined by a pharmacy store service area, a zip code area, a city, a county, and a statewide area; while the application of the claimed system and method is the same.

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		The system of variables and calculations demonstrated elsewhere in Goldenberg is of a mathematically very different nature with all the discussed shortcomings of statistics based approaches to the "noise" in data. Conclusion: The claims are different from the one in the reference.
	The third paragraph An algorithm for unusual public health status (or event) detection incorporating seasonally varying reference lines (see page 2, Tracking Grocery Data, paragraph 5), and calculating three structural components from the input data: • a daily deviation from the reference line (see Figure 3), • an n-days-cumulated-deviation, and (see Figure 3) • the change of the daily deviations in that area, (see Figure 3)	Quotation of reference U, page 2, Tracking Grocery Data, paragraph 5: " Our proposed detection system consists of several layers (A.G., G.S., and R.A.C., unpublished results). The first layer preprocesses the data by accounting for store level sales. The second layer puts the preprocessed data through a denoising filter. We use the discrete cosine transform (10, 11), which decomposes the series into cosine waves, and our filter retains only those that have a large magnitude. We chose the number of retained cosine waves to capture the main features of the series but also to avoid overfitting (A.G., G.S., and R.A.C., unpublished results)." Analysis: In reference U, (page 2, Tracking Grocery Data, paragraph 5) Goldenberg uses their 'unpublished results', and a wavelet transform, which is totally different from the approach in the Claim 1. In reference U, Figure 3, Goldenberg compares the sales data with the threshold which 'is in fact three standard deviations', as Goldenberg writes in Page 3, the second paragraph (above Figure 3), as a typical statistic approach. Goldenberg does not mention an n-days-cumulated-deviation in Figure 3, not in his paper either; Goldenberg does not mention the change of the daily deviation in

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		Figure 3, not his paper either. Conclusion: the claims are different from the one in the reference.
	The fourth paragraph A dynamic system model describing the categorized public health status by a set of state variables, and the change of the health status by the state transition, the input sets, the output sets, and the rule systems that govern them (see Page 3 paragraph 1)	Copy of Page 3, paragraph 1 of Reference U: " First, we decompose the denoised series into several "resolutions" by using a discrete (redundant) wavelet transform (ref. 12; cf. the continuous version of wavelets in ref. 13). Each resolution describes a different frequency of the series, but, unlike other transforms (e.g., the cosine and Fourier transform), it retains information on the *time* that each frequency is present. The resulting series for each resolution are more regular, and thus we use a simple autoregressive model (where the sales at time *t* are taken to be a weighted average of previous sales) for predicting each resolution separately. We then add the predictions to create the forecast of the next day sales. Fig. 2 shows the decomposition of the (preprocessed and denoised) series into five resolutions. For each resolution, we use an autoregressive model for forecasting the next point. Finally, we add the forecasts to obtain the next point in the series, i.e., Fig. 2 also includes the combined forecast of the next day (denoised) sales." Analysis: In Page 3, paragraph 1 of Reference U, Goldenberg describes 'a wavelet transform' and 'a simple autoregressive model'. His transform is a mathematical treatment/model of the problem which has no physical meaning nor describes a public health concept or content; his autoregressive model is for 'forecast of the next day sales'. The proposed patent matches each state variable with a public health concept/content, which is very unique to the proposed patent and not possible with "normal" modeling approaches. Goldenberg does not mention a model with a set of state variables, and does not model a change of the public health status by the state transition either.

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		Conclusion: the claims are different from the one in the reference.
	The fifth paragraph The rule system determines the state transitions for modeling the dynamic change of public health status through the analysis of information derived from OTC pharmaceutical sales in that area (see page 3, paragraph 1)	Copy of Page 3, paragraph 1 of Reference U: " First, we decompose the denoised series into several "resolutions" by using a discrete (redundant) wavelet transform (ref. 12; cf. the continuous version of wavelets in ref. 13). Each resolution describes a different frequency of the series, but, unlike other transforms (e.g., the cosine and Fourier transform), it retains information on the *time* that each frequency is present. The resulting series for each resolution are more regular, and thus we use a simple autoregressive model (where the sales at time *t* are taken to be a weighted average of previous sales) for predicting each resolution separately. We then add the predictions to create the forecast of the next day sales. Fig. 2 shows the decomposition of the (preprocessed and denoised) series into five resolutions. For each resolution, we use an autoregressive model for forecasting the next point. Finally, we add the forecasts to obtain the next point in the series, i.e., Fig. 2 also includes the combined forecast of the next day (denoised) sales." Analysis: In Page 3, paragraph 1 of Reference U, Goldenberg describes 'a wavelet transform' and 'a simple autoregressive model'. His transform is a mathematical treatment/model of the problem which has no physical meaning nor describes a public health concept or content; his autoregressive model is for 'forecast of the next day sales'. The proposed patent matches each state variable with a public health concept/content. Goldenberg's autoregressive model attempts to create the 'forecast of the next day sales'. Goldenberg does not mention the state transition, and the method he used here (simple autoregressive model) is a traditional statistical method, it is not a rule system, it is not for modeling the dynamic change of public health status either.

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		Conclusion: the claims are different from the one in the reference.
	The sixth paragraph A rule system combines the structural components incorporating the confidence supporting sets as the input variables (see page 3, paragraph 2 where the confidence arises by defining the input boundaries)	Quotation of paragraph 2 in page 3 of reference U: "The final layer of the detection system includes the computation of an upper threshold for the next day forecasts. This threshold is based on the forecast made in the previous step, plus a margin of error. When the actual next day sales become available, they are compared with the threshold. If they exceed the threshold, the system flags an alarm, indicating that the new daily sales are higher than expected. The threshold is based on the distribution of the differences between the forecasts and the real sales, and is in fact three standard deviations of the differences above the denoised series. This last step is based on a methodology used in statistical quality control, called control charts, where a process is monitored by using a chart that flags when a change occurs, while taking into account natural variation of the series (14). Fig. 3 illustrates the threshold for the cough OTC medication data. The threshold follows the series, creating a "security band," which, if exceeded, is an indication that the sales are higher than expected. For example, sales for 8_7_00 are higher than the prediction. They do not exceedthe threshold, however, and thus we do not take them to indicate an abnormal increase in sales." Analysis: In the Claim 1, three structured components are defined. The claim further describes a rule system combining the three components with confidence supporting sets (where the confidence supporting sets are defined by equation 7 - 9 in later section) as the input variables. The input variables are different from input data which is raw sale data. Also note, that confidence supporting set in the proposed method is not equivalent to 'three standard deviation'. In reference U, Goldenberg does not have those three components, no confidence supporting sets of those three components either. In reference U, page 3, paragraph 2, Goldenberg uses a threshold,

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		'this threshold is based on the forecast…, plus a margin of error', then to compare with the daily sales. (another example of analyst subjectivity in analysis, which the proposed method does not require) Conclusion: the claims are different from the one in the reference.
	The seventh paragraph A rule system maps the state history to the output variables (see Figure 5 where state history is determined by the amount of antibiotic being purchases).	Analysis: In reference U, Goldenberg has to create a simulation to evaluate their method, 'we devised a statistical simulation approach' (page 3, section of Evaluating the detection system, the first paragraph). And Goldenberg further assumes a medication sales pattern 'is a three-spike linearly increasing pattern', 'steadily over the first 3 days' (this is another example of analyst subjectivity). Goldenberg then 'measures the spike detection ratio (SDR, Goldenberg et al unpublished results)' (see page 3, paragraph 2 in reference U). Figure 5 shows the SDR. In Claim 1, this sentence describes a rule system that maps a set of state variables and their transition history into output variables; and the mapping is further defined by equation 12 in later section. Here the value of an output variable, as defined by equation 15, is the combination of three elements, the likelihood index of abnormality, the trend indicator and the potential impact index. Goldenberg's output is simply a ratio of spike detection. Goldenberg does not mention the state variables, no history of a set of state variables, no likelihood index, no trend indicator, no

Action #	Requirement/ Question / Advise by The Examiner	Analysis & Response from the Applicant
		potential impact index at all. Conclusion: The claims are different from the one in the reference.
27	As per Claim2, Goldenberg teaches the system wherein said measurement scheme includes the calculation of monthly (or weekly, or daily, or seasonally) averaged daily sales for the categorized OTC medicines as the base line, from the data in the past at the same place, which is one data set (base line) for supporting the rule system (see Figure 1).	Analysis: Figure 1 in Reference U, Goldenberg shows the OTC data noisy nature by plotting the raw data and the de-noising data. In reference U, Goldenberg does not mention the weekly or monthly averaged daily sales. The word 'weekly' are used twice total by Goldenberg in his paper: (1) one is about the data in ' medical and public health sources which are typically collected weekly or even less frequently.' (page 2 last paragraph), (2) the other is 'weekly effect showing higher sales during weekends'. No word 'monthly' ever used in Reference U by Goldenberg. Therefore, Goldenberg does not teach the system as described in claim 1 or 2. Conclusion: The claim is different from the one in the reference.
28	As per claim 3, Goldenberg teaches the system wherein said measurement scheme includes the calculation of the deviation of daily sales in the current-month from the base line, and it is measured in change of percentage at the same place, which is another data set (the first structural component) for supporting the rule	Analysis: In reference U, Figure 3, Goldenberg compares the sales data with the threshold which 'is in fact three standard deviations', as Goldenberg writes in Page 3, the second paragraph (above Figure 3), as a typical statistic approach. Goldenberg compares the raw sale data to their forecasts. His approach is different from the method defined in Claim 3. which

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
	system (see figure 3).	explicitly defined the baseline is derived directly from the historical data in that month at the same place, no forecast model is assumed, thus no subjectivity in model choice is necessary in the proposed patent. Conclusion: The claim is different from the one in the reference.
29	As per claim 4, Goldenberg further teaches the system wherein said measurement scheme includes the calculation of the n-days-cumulated-deviation, which is another data set (the second structural component) for supporting the rule system (see figure 3).	Analysis: In reference U, Figure 3, Goldenberg compares the sales data with the threshold which 'is in fact three standard deviations', as Goldenberg writes in Page 3, the second paragraph (above Figure 3), as a typical statistic approach. Goldenberg does not mention n-day-cumulated-deviation in his paper. No word 'cumulated' or the concept of 'cumulated-deviation' is used in his paper. Conclusion: The claim is different from the one in the reference.
30	As per claim 5, Goldenberg further teaches the system wherein said measurement scheme includes the calculation of the daily change of the deviation, which is another data set (the third structural component) for supporting the rule system (see figure 3).	In reference U, Figure 3, Goldenberg compares the sales data with the threshold which 'is in fact three standard deviations', as Goldenberg writes in Page 3, the second paragraph (above Figure 3), as a typical statistic approach. Goldenberg does not mention 'change of the deviation' at all. The word 'deviation' is used only once in his paper, where he describes his approach using in fact three standard deviations. In Claim 5, the calculation of the daily change of the deviation is

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		defined by equation 6. Conclusion: The claim is different from the one in the reference.
31	As per claim 7, Goldenberg further teaches the system wherein said dynamic model of the categorized public health status is defined by the system with a set of state variables and state transitions over the time dimension at a specified place, with which state transitions model the change of the categorized public health status (see page 2, Tracking Grocery Data section, paragraph 1 where the state variables and public health are tracked through antibiotic sales).	Quotation of Tracking Grocery Data section, paragraph 1 (of reference U): "**Tracking Grocery Data** Grocery and OTC medication sales have three main advantages for the detection of an outbreak: First, these datasets are typically very large and rich, including information on each purchased item and in many cases include customer information(e.g., address). They are also available on a more frequent scale, such as daily and even hourly basis, and do not include delays in reporting as compared with medical and public health sources which are typically collected weekly or even less frequently, and might contain delays. Second, the outbreak footprint would probably exist in these data earlier than in medical or public health data, because of self treatment that people usually pursue before seeking medical assistance. Third, although grocery and OTC sales do not measure illness directly, we might infer specific symptoms experienced by purchasers at a relatively early stage of the onset of the disease." Analysis: Goldenberg does not define the public health status by state variables, does not define the dynamic model that governs the stat transition which describes the dynamic process of public health status. Conclusion: The claim is different from the one in the reference.
32	As per claim 10, Goldenberg further teaches the system	Quotation of page 1 paragraph 3 in reference U: "We begin in the next section by providing background and a characterization of an outbreak of a bioagent, focusing on anthrax. Then we describe traditional data collected

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
	(1) wherein said input sets are the supporting sets for the state transition rule systems (see page 1 paragraph 3);	from medical and public health sources and their ability to detect attacks in a timely fashion, before turning to grocery data and the detection system that we developed. We also introduce a method for evaluating the detection system in the absence of a bioagent footprint in the data, and for tuning the system to the input data. We end with some observations on the usefulness of our approach." Analysis: Goldenberg does not mention state variables, no state transition, no supporting sets for the state transition either. Goldenberg mentioned 'input data'. However, Goldenberg's 'input data' is the simulated data, created with his assumptions including the time pattern and spike rising pattern. (see page 4, the second paragraph). In Claim 10, there are no assumptions for input sets. However, the applicant mended the sentence in Claim 10 to further clarify what is being mapped, please see the response to Action 23: "the structural components are mapped incorporating their confidence levels." Conclusion: The claim is different from the one in the reference.
	(2) it is mapped from the structural components incorporating the confidence levels (see figure 2 and page 3 paragraph 2 where statistical quality control creates	Analysis and response of (2) : Fig. 2 in reference U has two diagrams, and they are the results of Wavelet Transform for prediction. Goldenberg compares the sales data with the threshold which 'is in fact three standard deviations',

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
	confidence levels).	as Goldenberg writes in Page 3, paragraph 2 (above Figure 3), as a typical statistic approach, not in state pace, no rule system either. In Claim 10, the approach is defined in state space, and claim 10 further describes the input sets are supporting sets for state transition rule system, and the input sets are function of the confidence level for each structural component. Equation 11 defines the mathematical relation of them. Conclusion: The claim is different from the one in the reference.
33	(1) As per claim 11, Goldenberg further teaches the system wherein said input sets are the supporting sets for the state transition rule systems (see page 1 paragraph 3); (2) it is mapped from the structural components incorporating the confidence	Analysis of (1) : Quotation of page 1 paragraph 3 in reference U: "We begin in the next section by providing background and a characterization of an outbreak of a bioagent, focusing on anthrax. Then we describe traditional data collected from medical and public health sources and their ability to detect attacks in a timely fashion, before turning to grocery data and the detection system that we developed. We also introduce a method for evaluating the detection system in the absence of a bioagent footprint in the data, and for tuning the system to the input data. We end with some observations on the usefulness of our approach." Response: Goldenberg does not mention input sets, not state transition either; not in page 1 paragraph 3, not in his paper at all. Conclusion of (1): The claim is different from the one in the reference. Analysis of (2) In reference U, Figure 3, Goldenberg compares the sales data

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
	levels (see figure 3 and page 3 paragraph 2 where statistical quality control creates confidence levels), (3) where the confidence levels are derived from the historical data sets (see page 3 paragraph 2 where the historical data set are the natural variation), (4) and the confidence supporting sets are found from the cumulated distribution functions with the specified confidence levels (see figure 3 and page 3 paragraph 2 where the specified confidence level is the security band).	with the threshold which 'is in fact three standard deviations', as Goldenberg writes in Page 3, paragraph 2 (above Figure 3), as a typical statistic approach. Response to (2) : Goldenberg does not mention the structural components as described in Claim 11, he does not mention how to incorporate their confidence level either. Analysis of (3) and (4): See response to action 26 for Quotation of paragraph 2 in page 3 of reference U. Analysis: Goldenberg has a totally different approach. He computes the forecasts, compares the raw sale data to the forecasts plus a margin of error. The historical data set are not the natural variation, although the historical data sets (and the structural components) contain the natural variation. Goldenberg tries to use wavelet transform to model the natural variation. Goldenberg's security band is "in fact three standard deviations of the differences above the denoised series" as he wrote. In contrast, in Claim 11, the confidence supporting sets are functions (defined by equation 7, 8 and 9) which are a significant improvement and new invention, eliminating the analyst's subjectivity in choice of thresholds, and assumptions commonly found in his simulated input data. Response to (3) and (4): The applicant can not find how Goldenberg teaches the same

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		system.
34	As per claim 12, Goldenberg further teaches the system wherein said output sets are a set of vectors (see page 3 paragraph 1 where the vectors are resolutions), each with three values: likelihood, trend indicator, and impact indicator, where the output sets are mapped from the state history at the study place (see figure 2 where the values are related to their normalized counts).	Quotation of page 3 paragraph 1 in reference U: "First, we decompose the denoised series into several ''resolutions'' by using a discrete (redundant) wavelet transform (ref. 12; cf. the continuous version of wavelets in ref. 13). Each resolution describes a different frequency of the series, but, unlike other transforms (e.g., the cosine and Fourier transform), it retains information on the *time* that each frequency is present. The resulting series for each resolution are more regular, and thus we use a simple autoregressive model (where the sales at time *t* aretaken to be a weighted average of previous sales) for predictingeach resolution separately. We then add the predictions to create the forecast of the next day sales. Fig. 2 shows the decomposition of the (preprocessed and denoised) series into five resolutions. For each resolution, we use an autoregressive model for forecasting the next point. Finally, we add the forecasts to obtain the next point in the series, i.e., Fig. 2 also includes the combined forecast of the next day (denoised) sales." Analysis: Goldenberg uses wavelet transformations (not state variables) and gets several "resolutions". He does not say what wavelet function is used, and the 'resolution' is the results of transform in frequency, it has no physical meaning and no public health contents either. In the claims, including claim 12, in contrast to the reference U, state variables and a dynamic model of state transitions are used; Claim 12 describes how the history of the state variables are mapped into the output sets, and each output variable contains 3 elements, with public health meaning: likelihood, trend indicator, and impact indicator, which Goldenberg does not mention any of them (not in Fig 2 either) nor are any equivalent measures offered.

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		Conclusion: The claim is different from the one in the reference.
35	As per claim 13, Goldenberg further teaches the system wherein said rule system that governs the state transitions is the system with sets of logical rules, which evaluate both the logical and numerical functions to determine the system states (see page 2 paragraph 10 where the methodology encompasses two-stage prediction system with both logical and numerical functions).	Quotation of page 2 paragraph 10 in reference U: " The third layer of the system forecasts the next day sales given all of the previous sales. Although the data are now denoised, simple time-series models (e.g., autoregressive moving average models) do not perform well because of the non-stationarity of the series, i.e., the changes in their behavior over time cannot be characterized by simple time-series models. Instead, we use a two-stage prediction method suitable for non-stationary data that can be easily automated and yields more accurate predictions." Analysis: Goldenberg does not say what his two-stage prediction is; he does not mention the state variables, not state transitions either in his paper. And no rule system governs the state transition in his paper. Therefore, the claim is different from the one in the reference.
36	As per claim 14, Goldenberg further teaches the system wherein said rule system that processes the structural components is a rule system with both logical and numerical functions mapping the structural components to supporting sets (see figure 2).	Analysis: Goldenberg uses wavelet transform, not the state variables as in the claim 1- 15. In the proposed patent, the rule system processes state variables, with both logical and numerical functions. Figure 2 shows his wavelet transform results. There is no word of 'rule' or word of 'logic' is used in reference U. Therefore, the claim is different from the one in the reference.
37	As per claim 15, Goldenberg further	Analysis: Goldenberg uses wavelet transform and auto-regression

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
	teaches the system wherein said rule system that maps the state history to the output variables is a rule system with both logical and numerical functions mapping the state variables to the output variables which are described in Claim 12 (see figure 2 where the output variables are the prediction).	to forecast the sales; while the claims use state variable and dynamic model of state transitions, a rule system maps the state history into the output variables, each output variable has 3 elements as describes in Claim 12. Goldenberg does not use state variables, no rule system, nor any of those three elements. See response to Action 34. Therefore, the claim is different from the one in the reference.
38	Quotation of 35. U.S.C. 103(a)	Noted.
39	Claim 8 is rejected because of reference U and reference V on the attached.	See response to action 40.
40	(1) As per claim 8, Goldenberg teaches the apparatus of claim 7.... Armstrong teaches the apparatus, (see page 7 Measure paragraph 2)	See response to action 31. Goldenberg does not define the public health status by state variables, does not define the dynamic model that governs the state transitions which describes the dynamic process of public health status. No word of 'state' or 'dynamic model' is used in Reference U. Thus, the applicant can not agree 'Goldenberg teaches the apparatus of claim 7'.
40	(2) Armstrong teaches the apparatus, (see page 7 Measure paragraph 2)	Quotation of page 7 Measure paragraph 2 in reference V by Armstrong et al: "**Measures.** Parameters for measuring the importance of a health-related event—and therefore the public health surveillance system with which it is monitored---can include (7) • indices of frequency (e.g., the total number of cases and/or deaths; incidence rates, prevalence, and/or mortality rates); and summary measures of population health status (e.g., quality-adjusted life years [QALYS]); • indices of severity (e.g., bed-disability days, case-fatality ratio, and hospitalization rates and/or disability rates);

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		• disparities or inequities associated with the health-related event; • costs associated with the health-related event; • preventability (*10*); • potential clinical course in the absence of an intervention (e.g., vaccinations) (*11,12*); and • public interest. Analysis: The content described in the application Claim 1- 15 is 'a dynamic process of categorized public health status in a community (such as flu, or gastrointestinal diseases) at a place in a specified time, the process changes in hourly and daily, it is totally different from the above reference which measure by years, disability, and cost etc. Response: Applicant can not see how reference V defined the dynamic process of categorized public health status in a community with set of state variables, as healthy status, critical status, starting-unusual status, upward-trend-unusual status, peak-unusual status, downward-trend status, and ending-unusual status. The cited reference is delineating the administrative approach to public health status, not introducing a model by which the health status of a population can be described automatically and consistently. The state variables used in the proposed method and the health status indicators used in the cited sources are of a very different nature. One key difference is that the cited indicators do not relate to the field of early detection of public health events.
40	(3) Armstrong page 3, Summary paragraph 1	Quotation of reference V, page 3, Summary paragraph 1

Action #	Requirement / Question / Advise by The Examiner	Analysis & Response from the Applicant
		Summary *"The purpose of evaluating public health surveillance systems is to ensure that problems of public health importance are being monitored efficiently and effectively. CDC's* Guidelines for Evaluating Surveillance Systems *are being updated to address the need for a) the integration of surveillance and health information systems, b) the establishment of data standards, c) the electronic exchange of health data, and d) changes in the objectives of public health surveillance to facilitate the response of public health to emerging health threats (e.g., new diseases). This report provides updated guidelines for evaluating surveillance systems based on CDC's* Framework for Program Evaluation in Public Health, *research and discussion of concerns related to public health surveillance systems, and comments received from the public health community. The guidelines in this report describe many tasks and related activities that can be applied to public health surveillance systems."* Response: The applicant fully supports the above Summary, and systematically developed a new method (as detailed in claims 1-15). This method has been implemented in computer systems to improve the public health surveillance in real world. While the success of that implementation system can be assessed through the guidelines proposed in the cited source, the cited source does not disclose or cover the methods introduced by the application in any manner.

A Biointelligence System for Identifying Potential Disease Outbreaks

Monitoring Over-the-Counter Pharmaceutical Sales Data as an Indicator of Changes in Public Health Status

BY XIAOHUI ZHANG, RENO FIEDLER, AND MICHAEL POPOVICH

It can reasonably be expected that consumer spending patterns, such as purchases of over-the-counter pharmaceuticals, hold a strong relationship to the public health status, possibly indicating status changes earlier than traditional public health information systems. In the light of events over the last two years, such as the anthrax scare, the SARS outbreak, and other public health events, over-the-counter (OTC) data collection has received significant attention. Most OTC projects focus on data collection and basic statistical analysis. However, many questions have yet to be answered. For example: How does the OTC sales data link to the public health status and how can this be described in a rigorous framework? How can potential disease outbreaks be detected within the noise of real-world OTC data? Syndromic surveillance has come to the attention of public health only recently, so how can its results be integrated and validated within the existing body of public health knowledge? Overall, there exists no systematic framework for the potentially automated collection and analysis of OTC data for public health management consumption. This area remains largely unexplored. Small sampling sizes, significant fluctuations within OTC sales data, and the lack of evidentiary information confirming it make OTC sales surveillance systems challenging and difficult. These difficulties are magnified in that a biointelligence system (BIS) must become a component of the established public health information system infrastructure, while requiring an end-to-end implementation of advanced information technologies and a near real-time execution.

The development of public health surveillance systems requires multidisciplinary knowledge and advanced technologies. Halperin and Baker [1] provided an excellent summary on public health surveillance systems, and the Centers for Disease Control and Prevention [2], [3] developed new guidelines and recommendations for the evaluation of public health surveillance systems.

"The lessons learnt from the events following September 11, 2001, and the subsequent Anthrax attacks have proven that new and innovative technologies resources are absolutely necessary to ensure the nation is fully prepared" [4]. To that end, Scientific Technologies Corp. has developed the NH Pharmaceutical Sales Surveillance (NHPSS) for the New Hampshire Department of Health and Human Services (NH DHHS). The NHPSS was developed as a distributed information system. In addition to the database server, enterprise application servers and the Web-browser-based user interface, its architecture features knowledge-base technology, a new dynamic system model with rule systems, and automated data analysis in supporting public health surveillance. Internet mapping was also embedded in the system to provide for spatial analysis. Since its pilot application, started in December of 2002 in the Bureau of Communicable Disease Control and Surveillance, NH DHHS, the NHPSS has assisted NH DHHS in successful detections of gastrointestinal and respiratory events.

This article first introduces the methodology and the technology in the NHPSS development. Next, the system functionalities are introduced. Then, a new dynamic system model with a rule system for public health surveillance is described in detail. Finally its application and preliminary results are summarized.

Methodologies in Development of NHPSS

A Multidisciplinary Development

The multidisciplinary development of NHPSS draws from five system domains, going beyond real-time data compilation:

1) information technology (connectivity);
2) data models (adequate spatial-temporal dimensions);
3) knowledge-base (domain knowledge and data derived knowledge);
4) analytical methods (dynamic model and algorithms in automated processing);
5) enterprise applications.

Each of these domains plays an important role, and a systematic integration has been achieved in NHPSS. This article focuses on the data model, knowledge-base technology, and analytical methods developed in the NHPSS.

Data Processing and Information Organization

Massive OTC daily sales data are collected at the pharmacy

 0739-5175/04/$20.00©2004IEEE

stores. These sales can be categorized into medications treating respiratory-related syndromes, gastrointestinal-related syndromes, allergy syndromes, etc., according to their active ingredients. The categorized sales data at a store reflect the public health status, in this category, around that geographical area. The sales amount is furthermore impacted by the population in that area and the convenience to access the service (both hospital service and pharmacy service). After the purchase, a customer can take the medicine for several days. These factors and the spatial and temporal variations in the OTC sales have been quantitatively reflected in the NHPSS' data model and analytical process. To identify a potential outbreak, it is necessary to establish a set of reference lines for OTC sales. To that end, a measurement scheme has to be defined first. In the NHPSS, the geographical units are defined as the store service area, zip code area, city, and statewide area. In each geographical unit, the underlying population data (possibly with the age groups) can be abstracted from census data. A store's service area is derived from the driving distance between the store and its potential customers' homes. Time units used are daily, weekly, and monthly. Data processing and knowledge acquisition will be discussed in the following section.

From Raw Data to Spatial Data Warehouse

After replication of the daily OTC sales data, the raw data are automatically processed along spatial and temporal dimensions. In conjunction with a rule base, basic statistical methods have been applied here to derive the reference lines. The developed system requires a minimum of one-month historical data, while it is recommended that more than one year of data is available to improve the confidence set. Let x be the sales amount of a categorized medicine for a time unit (e.g., daily). This amount will be compared to reference lines (e.g. monthly) derived from historical sales records of the prior m-years, the specified category syndrome, and the geographical unit. The reference lines include a base line representing the regular daily amount and upper reference lines incorporating the confidence levels. Since the reference lines are computed periodically at each geographical level, the seasonal variations are maintained and the spatial characteristics are captured.

The rule system, integrated with the data warehouse approach, adapts the automated data processing and handles the exceptions. Two possible special cases have been considered: a) an epidemic outbreak was recorded in the history of this place, and b) there may be less than one year of historical data. The data warehouse organizes the seasonal varying reference lines at each geographical level.

It is worth noticing that a GIS tool was also integrated with the developed spatial data warehouse. The GIS derives and organizes the spatial background information, which includes the population with age groups. It also performs the spatial analysis.

First, a measurement scheme was defined to quantitatively and qualitatively evaluate the deviation of the incoming daily data from the reference line (to identify the possible abnormality) at each place. Next, a set of algorithms for a structural component analysis has been developed. The incoming OTC data are transformed into the structural components. A mapping of the structural components into the dynamic system model describes the change of public health status and identifies the possible unusual events.

A Dynamic Model for The Public Health Status

The dynamic process model in the state-space form was systematically formalized by Kalman, Falb, and Arbib (1968) [5]. Rosenbrock (1970) [6] expended multivariate systems in a state-space form. Since then, mathematical system theory, modern control engineering, and computer technology have enabled extensive successes in several industries. However, nonlinear systems in a state-space form are much less well understood. In public health, Castillo-Chavez, et al. 2002 [7] state that "the basic epidemiological equations are sufficiently nonlinear." The dynamic system model in state space as a means to describe this complexity has not been mentioned, as almost nothing is known about them in the public health context. This contrasts strongly with the commonly deployed intensive statistical approaches and stochastic epidemic modeling. A knowledge base with rule system can significantly improve decision-making support, because a large class of nonlinear functions can be described there, and a priori knowledge can be formulized from domain experts or derived from the data. Our effort has been to develop an integrated system that would embrace dynamic system theory, statistical methods, and a rule system with knowledge-base techniques as a unified tool. It is oriented toward problem solving in syndromic surveillance, but it is a biointelligence framework with many possible generalizations and implementations. The developed system can be implemented in a relatively short time, as demonstrated in the case of NHPSS.

Figure 1 shows the defined states and state transition diagram. The dynamic change of the public health status is modeled here in a new state-space form. This state-space form differs from the conventional state space approaches in that here the state transition, input mapping, and output mapping are governed by the rule system, while the conventional state-space form uses crisp algebra or linear algebra in most cases. With a state-space notation, at a specified place, at a specific time, a categorized health status is one of the following: healthy status (S_h), critical status (S_c), starting-unusual status (S_s), upward-trend-unusual status

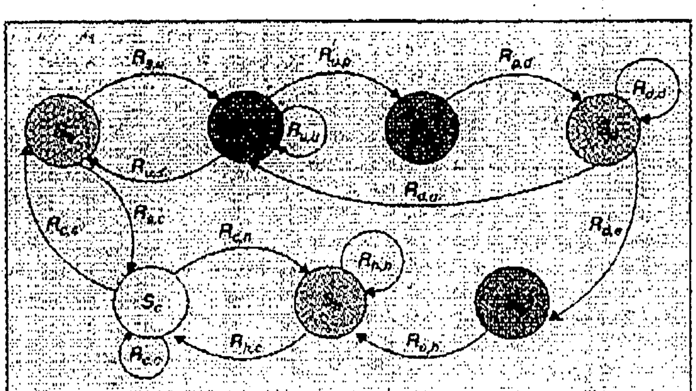

Fig. 1. The states and state transition diagram of the public health status.

Our effort has been to develop an integrated system that would embrace dynamic system theory, statistical methods, and a rule system with knowledge-base techniques as a unified tool.

(S_u), peak-unusual status (S_p), downward-trend-unusual status (S_d), and ending-unusual status (S_e). The state transitions over time reflect the dynamic change of the public health status.

The state space S is defined with its state variables $\{S: S_h, S_c, S_r, S_u, S_p, S_d, S_e\}$. A validated state transition from state $S_i(k)$ to state $S_j(k+1)$ is determined by the rule system that operates in relational algebra on its supporting set $X_i(k)$. The validated transition from state $S_i(k)$ to state $S_j(k+1)$ is determined by a rule base $R_{i,j}$, which evaluates the inputs $X_i(k)$ at state $S_i(k)$.

$$S_j(k+1) \Leftarrow \{S_i(k) \otimes R_{i,j} \otimes X_i(k),\ S_i(k) = k_w w_i(k)\} \quad (1)$$

$$X_i(k) \Leftarrow B_{i,m}(\alpha(k), \beta(k), \delta(k)) \otimes \begin{bmatrix} d_{i,m}(k) \\ w_{i,m}(k) \\ v_{i,m}(k) \end{bmatrix} \quad (2)$$

$$Y_i(k) \Leftarrow H_{i,n}\{\gamma_0(k), \gamma_1(k), \ldots, \gamma_n(K)\} \otimes \begin{bmatrix} S(k) \\ S(k-1) \\ \cdots \\ S(k-n) \end{bmatrix} \otimes G_i(k)\} \quad (3)$$

where

$$X \subset \bigcup_{q=1}^{Q} \mathrm{supp}(Q) \quad (4)$$

$$\mathrm{supp}(q) = \mathrm{supp}(X_{q,1}) \times \mathrm{supp}(X_{q,2}) \times \mathrm{supp}(X_{q,3}) \quad (5)$$

$$Y_i = (L_{i,h}, T_{i,h}, P_{i,h}). \quad (6)$$

In the defined system equations, there are three transformed components, $\{d_{i,m}(k), w_{i,m}(k), v_{i,m}(k)\}$, which are derived from the incoming raw data and then mapped into the supporting set. As time advances (for example, a time unit can be daily), the state transition from state $S_i(k)$ to state $S_j(k+1)$ is determined by the rule system $R_{i,j}$, which evaluates the supporting set $X_i(k)$, as shown in (1), where $\otimes$ stands for the inference operation, or a rule system operation, which can be logical operations or algebra operations or a hybrid. Equation (1) also defines the quantitative measurement for state $S_i(k)$ in that category at the specified place. The coefficient k_w can be defined by the user or it can be related to a threshold value obtained from the historical data set. The latter approach was taken in the NHPSS implementation to enable automated application of very granular rules without subjective input by experts.

Equation (2) describes that, at a state $S_i(k)$, there is the supporting set $X_i(k)$ with three structural components whose respective thresholds $\{\alpha(k), \beta(k), \delta(k)\}$ can be incorporated. The rule system $B_{i,m}$ maps the components into the supporting set $X(k)$.

Equation (3) describes the output mapping, which interprets the outputs from a set of states or a state history with the specified weight for the states by $\{\gamma_0(k), \gamma_1(k), \ldots, \gamma_n(k)\}$. In addition, the rule system combines the background information G_i, such as the environmental factors with the population demographics in the study area.

Equation (4) and (5) define the supporting system X as an additive combination of supporting sets, thus allowing inputs to be comprised of multiple data sources. In the NHPSS evaluation period, data sources included the OTC sales data, emergency department encounters, school closure events, and case count data.

Equation (6) defines that the value of an output is a combination of the likelihood index of abnormality ($L_{i,h}$), the trend indicator ($T_{i,h}$), and the potential impact index ($P_{i,k}$). An exemplary set has been defined here as:

$\{L_{i,h}$: (low, medium, high)$\}$,
$\{T_{i,h}$: (stable, upward, downward)$\}$,
$\{P_{i,h}$: (minor, moderate, sinificant)$\}$.

Consider the sample case where $Y_i = (L_{i,2}, T_{i,2}, P_{i,3})$. This case represents a medium likelihood abnormality, with upward trend status, and possible significant potential impact. In reality, this situation might require specified extensive management.

Knowledge-Base Technique and Rule Systems

A knowledge base could derive its information from data sets. Fensel and Studer (1999) [8] provide a comprehensive description on the application of knowledge acquisition and management. In NHPSS, a knowledge base compiles the incoming raw data into the designed forms. Next, it derives the relational facts, temporal characteristics, and regional patterns. Data processing methods include statistical analysis over space and time. The knowledge base organizes the information such that queries or evaluations posed to the knowledge base can be answered by means of an inference-based query-then-answering operation, or alternatively, an automated operation on evaluation and responses. During the development of NHPSS, GIS was integrated with the knowledge base. With knowledge-base technology, large data sets are processed along temporal and spatial dimensions. Information is derived to characterize the spatial distribution over time. Then the spatial data warehouse organizes the knowledge and its derivative information hierarchically by spatial areas. Figure 2 illustrates a spatiotemporal knowledge acquisition by deriving the seasonally varying reference lines, the trends, the extreme values, and the clusters for the OTC

sales in geographical dimensions for the categorized disease Syndromes in a local area incorporating regional and statewide information.

A rule system is a set of rules, arguments, constraints, relations, and responses. A rule can be numerical, logical, or both. A hybrid rule system consists of both explicit functions and logical rules. Bardossy and Duckstein (1995) [9] have an excellent introduction to rule systems. The rule system in NHPSS is a hybrid rule system that was developed for automated operations of the BIS. The rule system is implemented with a set of decision matrices. The developed rule system consists of sets of logical rules. Combining statistical analysis and epidemiology knowledge, the rule system evaluates the decision matrices and produces the responses. Examples are the comparison of the incoming data with the set of references. Differences are then quantitatively and qualitatively evaluated with respect to the space-time dimensions. The abnormalities of the OTC medicine sales are identified and assessed using relational algebra, relational calculus, and classical calculus. Figure 3 illustrates the rule system approach. Figure 4 shows the integration of the spatial data warehouse knowledge-base techniques with rule system to support the automated analysis and reporting.

Implementation of NHPSS

NHPSS was implemented as a distributed information system. Figure 5 shows the NHPSS system structure. Daily OTC pharmaceutical sales data are collected at each store, recorded at the pharmacy chain headquarters, and transmitted to NH DHHS. These data are replicated in data servers at the state public health department. The developed BIS data warehouse organizes data along logical dimensions. Next, automated data processing occurs in the application servers. Finally, analyses, reports, and alerts (if necessary) are generated to assist the decision making process of public health management. The user interface is Internet browser-based. With secured access, users can browse the data, search the reports and maps, and review the results of the trend analyses and unusual event detection methods as created by the built-in rule-base.

GIS plays a key role in the BIS spatial data warehouse with knowledge acquisition as was introduced above. It also performs the spatial analysis, such as abnormality analyses with scanning and ranking. Furthermore, GIS supports the outputs mapping for risk assessment as well as provides comprehensive reports with possible alerts. Figure 6 depicts the integration of GIS for knowledge acquisition and OTC analysis in NHPSS.

The main syndromic surveillance functions of the NHPSS BIS implementation can be summarized as:

(1) Automated data capture of OTC pharmaceutical sales data:

- Approximately 300 different pharmaceutical items are currently categorized into Gastrointestinal Diseases and Respiratory Illnesses.
- The measurement unit of disease indicators can be the daily number of sold packages (NHPSS), or the amount of sold active ingredient.

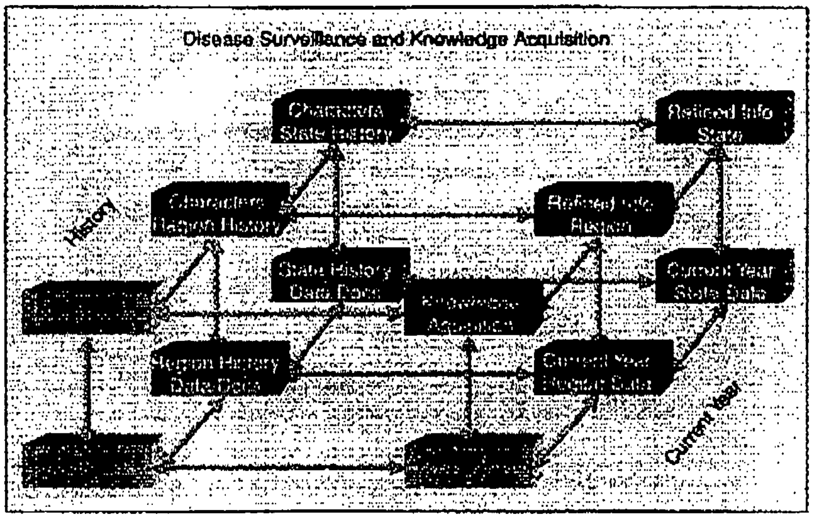

Fig. 2. Spatiotemporal knowledge acquisition scheme in NHPSS.

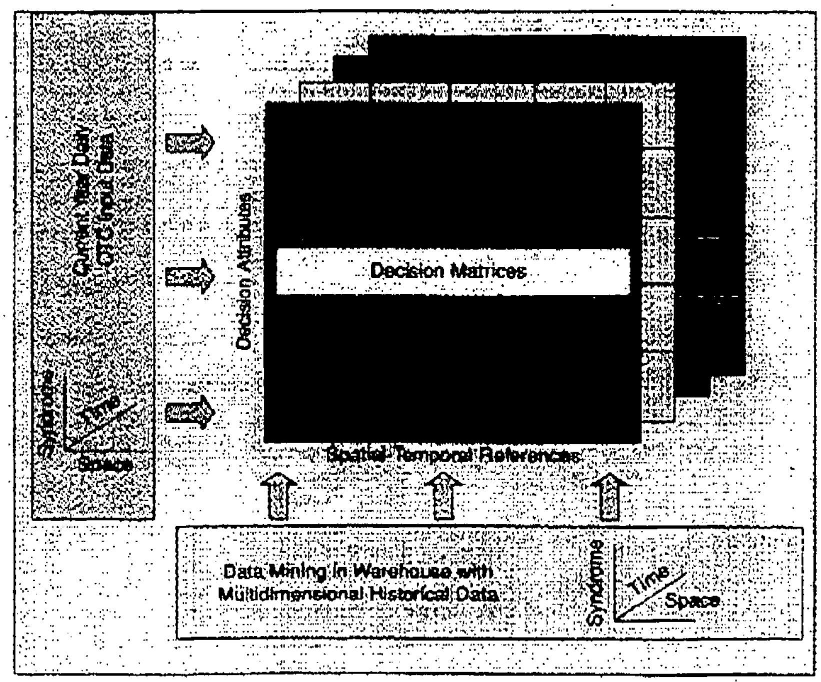

Fig. 3. Illustration of the rule system in NHPSS.

(2) Automated processing for data-derived reference lines (from historical data):
- ➤ Central Line: monthly-averaged (or weekly-averaged) daily sales
- ➤ Control lines: Min., Max., N-sigma lines and Confidence Interval Upper Limits.

(3) Analysis and Reporting in Time Dimensions:
- ➤ Detailed or aggregated reporting in daily/weekly/monthly for the selected place, with capability of comparison to the historical data.

(4) Analysis and Reporting in Geographical Areas:
- ➤ Map display with alerting capability for the specified time and disease indicators.
- ➤ Pinpoint the unusual areas.

(5) Rule-based Trend Analysis and Event Detections:
- ➤ detection of an unusual single-point-value event by comparison to the control lines
- ➤ detection of clusters and early warning of cluster-drifting
- ➤ detection and early warning of weekly average shifting
- ➤ detection and early warning of potential trend shifting
- ➤ detection of starting date, peak, and ending date of an event.

Figure 7 shows NHPSS hierarchical decision support, step by step, from a time series alert at the state level to pinpointing the unusual local areas and its detailed reports.

Pilot Application of NHPSS

The pilot application of the NHPSS BIS started in December 2002 at the Bureau of Communicable Disease Control and Surveillance (BCDCS), NH DHHS. Pharmacy stores in 23 cities reported daily OTC sales for a select set of pharmaceuticals to NH DHHS. The participating pharmacy stores represent approximately 10% of all stores in NH statewide and about 30% in the major cities. Since then, NHPSS has successfully supported BCDCS in detecting a large-scale gastrointestinal disease outbreak at the end of 2002 at both state and local levels, and a major influenza outbreak in February of 2003. The NHPSS output was compared to hospi-

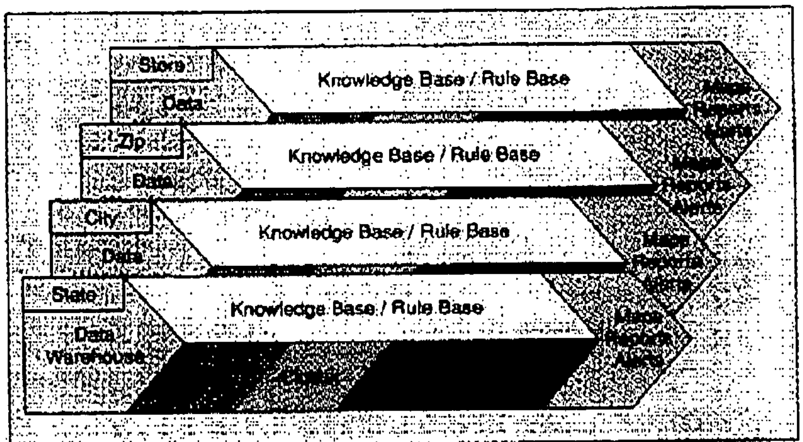

Fig. 4. Integration of spatial data warehouse, knowledge base, and rule system in NHPSS.

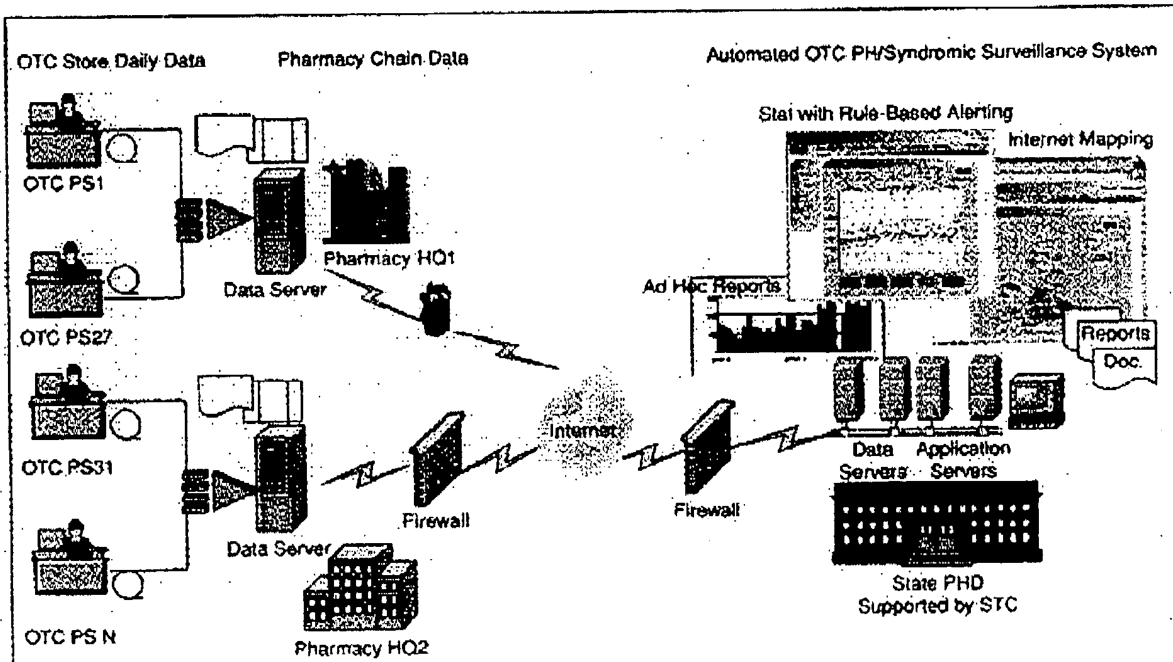

Fig. 5. Diagram of the NHPSS system structure.

tal ER counts, case counts, and laboratory information, and some school closure events. All data sources were at the same time period in the same area. The hospital ER data covers about 60% of total ER visiting in New Hampshire statewide. For the gastrointestinal disease outbreak, the NHPSS triggered an alert four days prior to the recognition of the outbreak at the state level. In local areas, such as communities, it has provided alerts up to ten days early. Furthermore, for some cities, NHPSS has provided alerts up to 12 days prior to the closure of schools, combating the spread of the influenza outbreak. Equally important is that the spatial and temporal characteristics of the outbreaks can be reported by the embedded Internet GIS application. Figure 8 displays the analysis results of NHPSS outputs, with GIS tool scanning and ranking the abnormality in support of risk assessment. Several other less impactive localized and state-wide events have been detected and described by the NHPSS BIS.

Conclusions

Developing syndromic surveillance systems to support early warning of both natural and possibly intentional public health events has received significant attention in the past two years. This article has introduced a biointelligence system that integrates advanced information technology with dynamic system theory to support the early detection of potential disease outbreaks. A new dynamic model was developed in the state-space form to describe the change of public health status incorporating multiple sources of syndromic surveillance data. A measurement scheme for the spatially and temporally varying time series data was developed. Spatial data warehouse and knowledge-base techniques are integrated with automated data processing and automated knowledge acquisition. A companion rule system governs the state transitions and supports the event detection. The implemented system, NHPSS, has Internet GIS to support spatial analysis and decision-making, as well as a commercial-off-the-shelf Web-based reporting tool. The pilot application has yielded promising results. Beyond the merits of the developed system for the OTC sales surveillance, the framework has been generalized for multiple data sources. This system is still in the early stage, but the preliminary results already start to challenge the traditional methods in their own area of excellence or where they cannot be applied.

NHPSS has demonstrated that public health management can greatly benefit from early warnings for disease outbreaks through the implementation of automated syndromic surveillance. The largest remaining challenge is the cooperation of multiple organizations and the private sectors in sharing the information for the common goal while preserving confidentiality of proprietary interests such as market penetration by individual pharmacy chains.

Acknowledgments

The authors sincerely thank Dr. J. Greenblatt and Ms. S. Miller at New Hampshire DHHS, who had the vision to start this innovative project with us.

Xiaohui Zhang is a senior scientist at Scientific Technologies Corp. He graduated in 1997 with a doctoral degree in resource management and in 1993 with a master's degree in systems engineering, both from the University of Arizona, Tucson, Arizona. Prior to STC, Dr. Zhang was the chief scientist of (now former) Intersect Technologies Corp, where he developed new algorithms to improve the enterprise network management, and he was a senior researcher on high-performance modeling and simulations at the

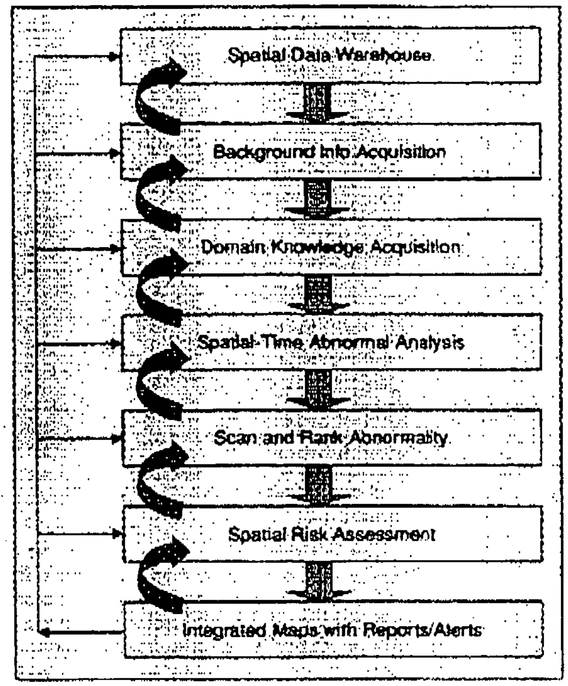

Fig. 6. Integration of GIS with knowledge base and spatial analysis.

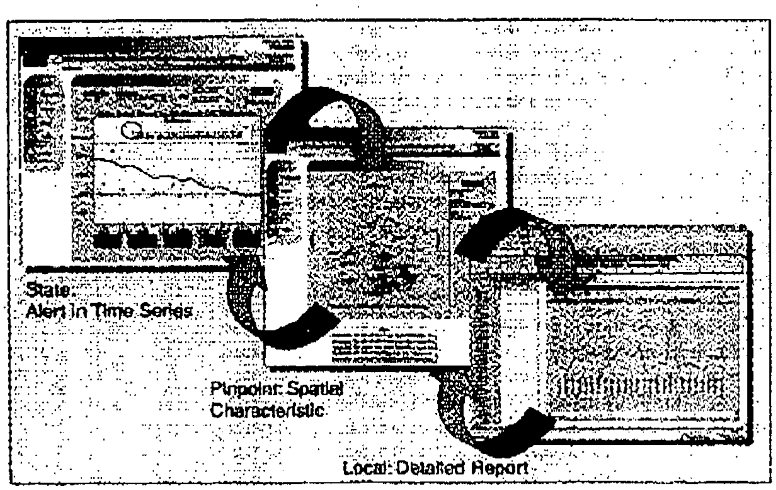

Fig. 7. NHPSS hierarchical decision support pinpointing unusual area and detailed report.

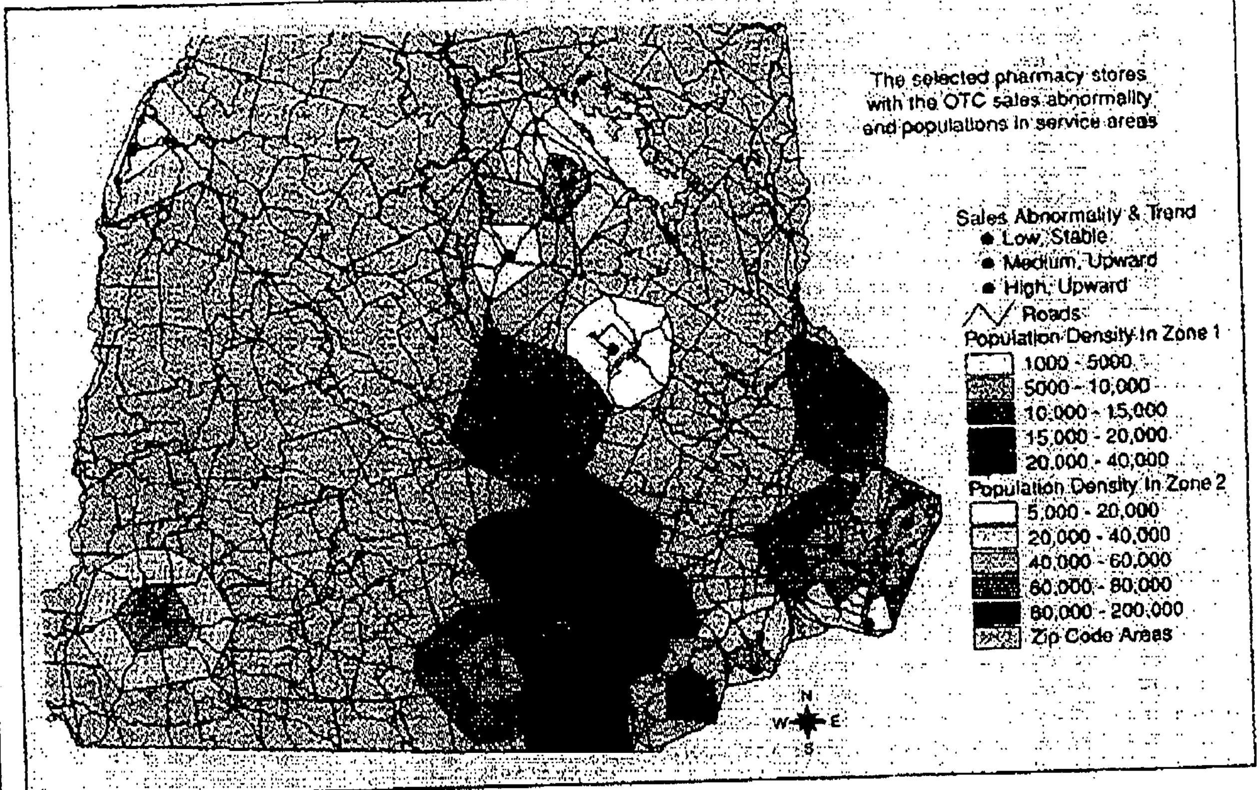

Fig. 8. An example of abnormality analysis to support risk assessment.

University of Arizona. He has 18 years of experience in modeling and simulation of complex systems. He has over 20 publications on decision-making techniques under uncertainty, optimization, artificial intelligence, modeling and simulation of complex systems, distributed simulation with remote sensing, and GIS applications.

Reno Fiedler is executive director of research and development at STC. He received his master's degree in geological engineering and geophysics from the University of Arizona in 1995. He has been a senior system architect in design and development of distributed information systems with GIS for public health and environmental applications. With ten years of international consulting experience, his expertise includes strategic planning, system design, and Internet GIS application development. He was an invited speaker at a national GIS conference. He has published on pattern cognition using fuzzy set theory, the use of fractals in a geo-engineering topics, and spatial statistics in geology.

Michael Popovich is founder and president of Scientific Technologies Corporation. He graduated in 1973 with a master's of science in systems engineering from the University of Arizona and with a B.S. in engineering mathematics and a minor in physics. He has received numerous awards, authored over 120 articles and papers, and is a regular invited guest speaker at national conferences. He founded Scientific Technologies Corporation in 1988, a leading-edge software company specializing in information technology planning and implementation. He has designed and applied methodologies to develop information technology strategic plans as well as analyze existing information systems. Since the early 1990s he has become one of the nation's leading specialists designing, building, and implementing public health information systems, most notably, immunization registries and communicable disease reporting systems.

Address for Correspondence: Michael Popovich, Scientific Technologies Corporation, 4400 E. Broadway Blvd, Tucson, AZ 85711 USA. Tel.: +1 520 202 3333. Fax: +1 520 202 3340. E-mail: michael_Popovich@stchome.com.

References

[1] W. Halperin and E.L. Baker, *Public Health Surveillance*. New York: Van Nostrand Reinhold, 1992.

[2] CDC 1998, "Guidelines for evaluating surveillance systems," *MMWR*, vol. 37, no. S-5, May 6, 1998.

[3] CDC 2001, "Updated guidelines for evaluating public health surveillance systems: Recommendations from the Guidelines Working Group," *MMWR*, vol. 50, no. RR-13, July 27, 2001.

[4] M.L. Popovich, J.M. Henderson, and J. Stinn, "Information technology in the age of emergency public health response," *IEEE Eng. Med. Biol.*, vol. 21, no. 5, Sept. 2002.

[5] R.E. Kalman, P.L. Falb, and M.A. Arbib, *Topics in Mathematical System Theory*. New York: McGraw-Hill, 1968.

[6] H.H. Rosenbrock, *State-Space and Multivariable Tehory*. London: Thomas Nelson and Sons LTD, 1970.

[7] C. Castillo-Chavez et al, *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction*. New York: Springer, 2002.

[8] D. Fensel and R. Studer, Eds., *Knowledge Acquisition, Modeling and Management*. 11th European Workshop, EKAW'99, Springer, 1999.

[9] A. Bardossy and L. Duckstein, *Fuzzy Rule-Based Modelling with Applications*. Boca Raton, FL: CRC Press Inc, 1995.